

<https://helda.helsinki.fi>

Stimulus duration has little effect on auditory, visual and audiovisual temporal order judgement

Tiippana, Kaisa

2018-05

Tiippana , K & Salmela , V R 2018 , ' Stimulus duration has little effect on auditory, visual and audiovisual temporal order judgement ' , Experimental Brain Research , vol. 236 , no. 5 , pp. 1273-1282 . <https://doi.org/10.1007/s00221-018-5218-2>

<http://hdl.handle.net/10138/310047>

<https://doi.org/10.1007/s00221-018-5218-2>

acceptedVersion

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.

Stimulus Duration Has Little Effect on Auditory, Visual and Audiovisual Temporal Order Judgement

Kaisa Tiippana and Viljami R. Salmela

Department of Psychology and Logopedics, Faculty of Medicine, University of Helsinki, Finland

Corresponding author: Dr. Kaisa Tiippana, kaisa.tiippana@helsinki.fi, tel. +358-503185721, complete
postal address: Department of Psychology and Logopedics, Faculty of Medicine, P.O. Box 21, 00014
University of Helsinki, Finland

ORCID 0000-0002-2305-8104

Abstract

Some classical studies on temporal order judgments (TOJ) suggested a single central process comparing stimulus onsets across modalities. The prevalent current view suggests that there is modality-specific timing estimation followed by a cross-modal stage. If the latter view is correct, TOJ's may vary depending on stimulus modality. Further, if TOJ is based only on onsets, stimulus duration should be irrelevant. In order to address these issues, we used both unisensory and multisensory stimuli to test whether unisensory duration processing influences cross-modal TOJ's. The stimuli were auditory noise bursts, visual squares, and their cross-modal combinations presented at 10, 40 and 500 ms durations, and various stimulus onset asynchronies. Psychometric functions were measured with an identical task in all conditions: On each trial, two stimuli were presented, one to the left, the other to the right of fixation. The participants judged which one started first. TOJ's were little affected by stimulus duration, implying that they are mainly determined by stimulus onsets. Throughout, the cross-modal just noticeable differences were larger than the unisensory ones. In accordance with the current view, our results suggest that cross-modal TOJ's require a comparison of timing after modality-specific estimations.

Keywords

audiovisual; cross-modal; duration; multisensory; temporal order judgment; time perception

Introduction

The assessment of the ability to detect cross-modal synchrony dates back to the origins of experimental psychology over two hundred years ago (as described e.g. by Mollon & Perkins, 1996). The precision of judgment and the relative timing of lights and sounds that is required for perceived synchrony have since been extensively studied (van Wassenhove, 2009). Despite this, several issues still remain unresolved, highlighting the complexity of the apparently simple question of when two stimuli appear to be synchronous.

In most studies investigating the perception of multisensory synchrony, only cross-modal conditions are measured, without any unisensory conditions. This is despite the well-known fact that the manipulation of various stimulus properties influences unisensory processing in general (e.g. intensity; Salmela & Laurinen, 2009, frequency; Patterson et al., 1982, complexity; Tiippana, Näsänen & Rovamo, 1994, duration; Borucki & Berg, 2017), which could in turn be reflected in cross-modal processing. For example, when stimulus duration (1 and 13 s) and blur level of audiovisual video excerpts of speech are manipulated, simultaneity judgments result in a wider window of perceived synchrony for short than long excerpts, but blur does not influence judgments (Eg and Behne, 2015). The synchrony window is also wider when the visual stimulus is dim than bright (Krueger Fister, Stevenson, Nidiffer, Barnett, and Wallace, 2016). However, how these manipulations affected unisensory synchrony judgments was not studied.

An influential study that includes temporal order judgment (TOJ) measurements for both unisensory and cross-modal stimuli is by Hirsh and Sherrick (1961), who claimed that the just noticeable differences (JND) are similar (in the order of 20 ms) for all conditions: auditory, visual, tactile and their combinations. It seems that in subsequent multisensory TOJ studies, unisensory performance has not been measured since the (often implicit) assumption has been that it does not differ from cross-modal performance. These findings gave rise to the idea that timing judgments are based on a central process that operates independently of sensory processes (e.g. Sternberg and Knoll, 1973).

However, some evidence exists that the precision of cross-modal TOJ's can be poorer than that of unisensory TOJ's. Zampini, Shore and Spence (2003) have suggested that Hirsh and Sherrick's (1961) results may represent an upper limit of cross-modal performance since the conditions differed for example in spatial disparity. Their own data had a trend for audiovisual (AV) JND's to be larger than unisensory ones. In support of this, two studies on the perceptual learning of temporal order found larger JND's for AV than auditory (A) or visual (V) judgments (Alais and Cass, 2010; Virsu, Oksanen-Hennah, Vedenpää, Jaatinen, and Lahti-Nuuttila, 2008). Synchrony judgments of temporally modulated stimuli have also shown similar results (Fujisaki and Nishida, 2005, 2010). Differences in timing judgments due to different sensory modalities speak against a single, central time processor

(Fujisaki, Kitazawa, and Nishida, 2012; Fujisaki and Nishida, 2005, 2010). Another line of evidence against a single time processor is the well-established finding that judgments of duration are influenced by stimulus modality, for example so that auditory stimuli are judged shorter than visual stimuli of equal physical duration (e.g. Grondin, 2014; Penney, Gibbon & Meck, 2000; Shi, Church & Meck, 2013).

In TOJ, it might easily be assumed that only stimulus onsets are compared. By varying stimulus duration, this assumption can be tested since stimulus duration is irrelevant and should not matter if only onsets are compared. Still, timing judgments may be affected by stimulus offset or duration. There are few previous studies investigating the effect of stimulus duration on audiovisual synchrony perception. Boenke, Deliano and Ohl (2009) found that the PSS shifted toward a common attractor value with increasing stimulus duration, but there was no effect on the JND. Kuling, van Eijk, Juola, and Kohlrausch (2012) used an audiovisual synchrony judgment task and found no effects when the duration was the same for auditory and visual stimuli. In contrast, for stimuli with different durations, a longer visual lead was required when the visual stimulus was long rather than short, and vice versa for auditory stimuli. These findings led the authors to postulate that “there might exist systematic unimodal differences between physical and perceived onsets for the experimental stimuli. Such relative within-modality shifts could quite well lead to quantitative shifts in intermodal timing parameters, like the PSS. Only when those unimodal effects are quantitatively accounted for can one correctly interpret the (remaining) temporal effects as reflecting true intermodal properties.” However, these studies included only cross-modal conditions. In a unisensory study, auditory JND's were independent of duration for short (10-40 ms) dichotic stimuli (Babkoff and Fostick, 2013). The perceived onset position of an auditory stimulus is affected by duration in some cases, but not in others (Schimmel and Kohlrausch, 2008). Similarly, in tasks to match the onset of two visual stimuli, the matches are sometimes influenced by duration and sometimes not (Jaśkowski, 1991; 1992; 1993). Overall, the duration effects remain rather variable (see Grondin, 2010, for a general review). Furthermore, the duration effects have not been studied together for unisensory and cross-modal stimulus conditions.

Finally, temporal judgments are known to show considerable individual variation (e.g. Boenke et al., 2009; Grabot & van Wassenhove, 2017; Mollon and Perkins, 1996; Stevenson, Zemtsov and Wallace, 2012). These may contribute to the mixed findings regarding duration effects reported previously. Therefore, we investigated both averaged TOJ behaviour to find out general patterns as well as individual differences contributing to the phenomena.

The aim of the current study was to address the question whether unisensory TOJ's are reflected in audiovisual TOJ's, and whether TOJ's are influenced by stimulus duration. We used the same left-right task of judging stimulus onsets, and the same stimulus durations (10, 40, 500 ms) in all

conditions. These parameters were selected to be very similar to those used by Boenke et al. (2009), with the aim to make their and our results more comparable. They also used a left-right task instead of the more commonly used cross-modal task of reporting which stimulus modality was perceived first. A spatial task avoids a bias to report one modality over the other. In the current study, it also enabled the use of the same task in all unisensory and audiovisual conditions. We had conditions where both stimuli had equal durations (as did Boenke et al., 2009), as well as conditions where the duration varied only for one stimulus while remaining constant for the other. We introduced the latter variation because Kuling et al. (2012) found that synchrony judgments were influenced by duration only when stimuli had unequal durations. Duration effects might become more prominent also for TOJ's in the latter case, i.e. when stimulus durations differ.

Three different patterns of results can arise. Firstly, if TOJ is determined by a single, central timing mechanism comparing stimulus onsets, duration should not have any effect, and JND's would be similar in all modalities. Secondly, if the processing in the central timing mechanism were not based solely on onsets, stimulus duration should influence TOJ's similarly for all modalities. Thirdly, if stimulus timing is estimated separately for each modality, the unisensory TOJ's might differ and also vary as a function of duration, and this could be reflected in AV TOJ's. For example, auditory JND's might decrease with duration, meanwhile visual JND's could increase with duration, resulting in audiovisual JND's which are independent of duration. Further, based on previous findings, we expected that unisensory JND's would be smaller than audiovisual JND's. This, together with any effects of stimulus duration that differ between modalities would give further support to the existence of multiple time processors.

Methods

Participants

In total 64 participants (average age 25.3 years, SD 7.3; 20 males; 5 left-handed) took part in the experiments. They reported normal vision and hearing. They were first year psychology students and the experiments were conducted as a part of a methods course in student laboratories at the University of Helsinki. They were aware of the objectives and the methods of the study, gave informed consent and received study credit for participation. This research was carried out in agreement with legal requirements and international norms (Declaration of Helsinki, 1964) and the research has been approved by the Ethical Review Board in Humanities and Social and Behavioural Sciences at the University of Helsinki.

Stimuli and equipment

The auditory stimuli were bursts of white noise with three durations: 10, 40 and 500 ms. The fade in/out time was 2 ms, which eliminated any audible clicks at on/offset. The intensity was 60 dB(A). The visual stimuli were flashes of white (100 cd/m²) 3 deg squares on a mid-gray background (50 cd/m²) with three durations: 10, 40 or 500 ms.

The experiments were conducted in two laboratories with equal equipment. The auditory stimuli were delivered via calibrated active speakers (Genelec 6010A). The visual stimuli were presented on a calibrated CRT-display (Sony CPD-G420; display area 20.9 x 15.8 deg; 100 Hz refresh rate). The stimuli were generated with Matlab 7.10 (Mathworks Inc, Natick, MA, US), using PsychPortAudio driver and Psycyhtoolbox 3.0.8 (Brainard, 1997; Kleiner, Brainard, and Pelli, 2007; Pelli, 1997). The timing of the stimuli was calibrated and the audiovisual precision of the system was approximately 1 ms. We measured the precision of our stimulus system using light sensitive photodiodes. The timing of response of the photodiode and the output of audio card was compared, and the error in timing was about a millisecond.

The participants sat at a distance of 80 cm from the display and used a chin-rest for steady head position. The speakers were on the left and right side of the display. The distance from the centre of the display to the centre of the speakers was 20 deg. The visual stimuli were presented on the left and right side of the display at 11 deg eccentricity.

Procedure

The effect of stimulus duration on temporal order judgment (TOJ) was measured in three conditions: auditory (A), visual (V) and audiovisual (AV). The point of subjective simultaneity (PSS) and discrimination thresholds, i.e. just noticeable differences (JND's), were measured with the method of constant stimuli.

A pair of stimuli was presented on each trial (two flashes, two sounds or one flash and one sound). The flashes were presented on the left or right side of the display and the sounds were presented from the left or right speaker (as in Boenke et al., 2009). The participants' task was to respond on which side the stimulus started first by pressing one of two keys on the keyboard: left/right. The participants were instructed to base the decision on the onset of the stimulus. Thus, the duration of the stimulus was completely irrelevant to the task. The next trial started randomly 500-1000 ms after the response, with the start times selected from a uniform distribution. Participants were instructed to fixate on a cross that was displayed on the center of the screen throughout.

To measure the psychometric functions (PF), the temporal asynchrony of the stimuli was varied. Eight stimulus onset asynchronies (SOA) were used. The SOA's for visual, auditory and audio-visual conditions were $\pm 20, 50, 80, 110, \pm 20, 60, 100, 140$ and $\pm 40, 120, 200, 280$ ms, respectively. The SOA's for the different conditions were determined on the basis of measurements done on the same course the preceding year, with different participants who measured unisensory and audiovisual TOJ PF's with various stimulus parameters, excluding duration manipulation. In the present experiment, each SOA was repeated 20 times and thus each condition consisted of 160 trials. The A and V conditions were measured once and the AV conditions were measured twice. Thus, each psychometric function for every participant was based on 160 trials in A and V conditions and on 320 trials in AV conditions.

In every condition (A, V, AV), the duration of the stimulus was varied (10, 40, 500 ms). Further, either both stimuli were of equal duration, or one varied in duration while the other stimulus had a constant duration of 10 ms. Stimulus durations and SOA's are demonstrated in Fig. 1.

Each participant measured 7 conditions of which 3 were audiovisual and 4 unisensory. The conditions were measured in separate blocks. The measurement order of the blocks was randomized. The participants were randomly divided into 4 groups (Table 1). Group 1 ($n=15$) had stimuli with equal durations, and the participants measured three auditory and audiovisual conditions and one visual condition (same duration, 3 auditory conditions). Group 2 ($n=15$) had stimuli with different durations, and the participants measured three auditory and audiovisual conditions and one visual condition (different duration, 3 auditory conditions). Group 3 ($n=19$) had stimuli with equal durations, and the participants measured three visual and audiovisual conditions and one auditory condition (same

duration, 3 visual conditions). Group 4 (n=15) had stimuli with different durations, and the participants measured three visual and audiovisual conditions and one auditory condition (different duration, 3 visual conditions).

Table 1. about here

Figure 1. about here

To ensure that the participants were familiar with the tasks and conditions, they practiced by measuring every condition once. These practice data were discarded from the analyses. After the practice, each participant proceeded to the actual measurements (10*160=1600 trials). The measurements were done on 3-5 days within a period of two weeks.

Data analysis

In the unisensory conditions, the psychometric function was obtained by calculating the number of “right side first” responses as a function of SOA, and a cumulative Gaussian function was fitted to the data using Matlab’s fit-function. In order to account the individual variability in the data, the fitted Gaussian had four free parameters: mean, standard deviation, guess rate and lapse rate (Wichmann and Hill, 2001). The guess rate allowed the psychometric function to start from values 0-0.15 and the lapse rate allowed the function to end to values 0.85-1. The 0.5 point of the psychometric function was defined as the PSS and the difference between the 0.75 point and the PSS was defined as the JND. Note that the PSS here should be close to zero; any deviation would indicate a bias to respond “left” (negative values) or “right” (positive values).

In cross-modal conditions, the analyses were done in two ways (following Zampini et al., 2003). First, they were done as in the unisensory conditions, by calculating the proportion of “right side stimulus first” responses. This way A, V, and AV results were comparable, expressed as “right first”. Second, the “left”/“right” responses were converted to correct/incorrect responses at each SOA. For the series where the visual stimulus was presented first, the number of correct responses was calculated (visual stimulus started first on the left and the participant responded “left” + visual stimulus started first on the right and the participant responded “right”). For the series where the auditory stimulus was presented first, the number of incorrect responses was calculated (auditory stimulus started first on the left and the participant responded “right” + auditory stimulus started first on the right and the participant responded “left”). The psychometric function showing the number of “visual stimulus first” responses as a function of SOA was obtained by plotting the former to positive SOA values, and

the latter to negative SOA values. The psychometric function expressed as “vision first” thus indicates the relative timing between auditory and visual stimuli, and is the one that is traditionally used in AV TOJ measurements. Its PSS indicates whether the auditory stimulus (negative values) or the visual stimulus (positive values) needs to lead in order for the cross-modal stimuli to appear simultaneous.

In every condition, psychometric functions were calculated individually for each participant (7x64=448 functions in total). If the fit of the function was bad and the coefficient of determination (R^2) was below 0.7, the function was discarded. In total, 31 out of 448 functions were rejected (6.9% of the data; see Fig. 2 for R^2 vs JND scatter plot). Finally, the values were averaged across participants separately for JND and PSS and subjected to statistical analyses.

The effects of presentation modality and stimulus duration on right first JND's were examined by conducting a two-way repeated-measures ANOVA with Modality (2 levels: unisensory, audiovisual) and Duration (3 levels: 10, 40, 500 ms) as factors for each group separately. When the interaction was significant in the two-way ANOVA, a simple repeated-measures ANOVA with Duration (3 levels: 10, 40, 500 ms) as a factor was conducted for each group and modality separately. Vision first JND's were analysed similarly to right first JND's.

For right first PSS's similar analyses were conducted. In all cases, right first PSS's only indicate a response bias for the left or right side. Instead, the AV vision first PSS's are informative since they indicate whether the auditory or visual stimulus has to lead for simultaneity perception. Since for vision first analyses the audiovisual values are not directly comparable with the right first unisensory analyses, a simple repeated-measures ANOVA with Duration (3 levels: 10, 40, 500 ms) as a factor was conducted for each group separately for AV vision first PSS's.

Greenhouse-Geisser corrections were used when appropriate. Pairwise comparisons were done using two-tailed paired t-tests. The significance level was taken to be $p < .05$. Statistical analyses were conducted with SPSS software (version 24.0.0.0).

Figure 2. about here

Results

Psychometric functions

The psychometric functions were of good quality, and the fits were quite good (Fig. 2). Still, the variation between individuals was large. We shall return to these points with the derived measures of JND and PSS below.

Just noticeable differences

The JND's for auditory, visual and audiovisual stimuli as a function of stimulus duration for Groups 1-4 are shown in Figure 3.

For Group 1 (same duration, 3 auditory conditions) and right first analysis, there was a significant main effect of Modality [$F(1,10)=9.57$, $p=.011$, $\eta_p^2=.489$] because audiovisual JND's were larger than auditory ones (Fig. 3A). The main effect of Duration was not significant [$F(2,20)=.488$, $p=.56$, $\eta_p^2=.044$], but the interaction was [$F(2,20)=7.72$, $p=.007$, $\eta_p^2=.436$]. A simple ANOVA for each modality showed that the effect of duration was significant for auditory stimuli [$F(2,28)=5.92$, $p=.007$, $\eta_p^2=.297$], which was due to a decrease of A JND's with duration. Pairwise comparisons showed that the JND was higher for 10 ($p=.030$) and 40 ($p=.020$) than 500 ms duration. There was no effect of duration for audiovisual stimuli [$F(2,20)=2.63$, $p=.13$, $\eta_p^2=.208$]. The analysis of the results expressed as vision first for AV provided the same significant results. There was a significant main effect of Modality [$F(1,10)=15.0$, $p=.0103$, $\eta_p^2=.599$] because AV JND's were larger than auditory ones. The main effect of duration was not significant [$F(2,20)=.016$, $p=.98$, $\eta_p^2=.002$], but the interaction was [$F(2,20)=4.58$, $p=.023$, $\eta_p^2=.314$]. A simple ANOVA showed no effect of duration for audiovisual stimuli [$F(2,20)=1.17$, $p=.32$, $\eta_p^2=.105$].

For Group 2 (different duration, 3 auditory conditions) and right first analysis, the main effect of Modality approached significance [$F(1,13)=3.75$, $p=.075$, $\eta_p^2=.224$] since audiovisual JND's were slightly larger than auditory ones, but it did not reach significance (Fig. 3B). The main effects of Duration [$F(2,26)=2.415$, $p=.14$, $\eta_p^2=.142$] and interaction [$F(2,26)=.175$, $p=.84$, $\eta_p^2=.013$] were not significant. The analysis of the results expressed as vision first for AV provided similar non-significant results. [Modality $F(1,13)=1.60$, $p=.23$, $\eta_p^2=.109$; Duration $F(2,26)=1.01$, $p=.38$, $\eta_p^2=.072$] interaction [$F(2,26)=.712$, $p=.50$, $\eta_p^2=.052$].

For Group 3 (same duration, 3 visual conditions) and right first analysis, there was significant main effects of Modality [$F(1,13)=36.1$, $p<.001$, $\eta_p^2=.735$] and Duration [$F(2,26)=9.64$, $p=.001$, $\eta_p^2=.426$], but the interaction was not [$F(2,26)=.195$, $p=.82$, $\eta_p^2=.0150$]. This was because audiovisual JND's were larger than visual ones, and both increased at 500 ms (Fig. 3C). Pairwise comparisons showed that the V JND was lower for 10 than 500 ms ($p=.036$) and AV JND was lower for 40 than 500 ms duration ($p=.005$). The analysis of the results expressed as vision first for AV provided the same significant results. There was a significant main effect of Modality [$F(1,13)=24.9$, $p<.001$, $\eta_p^2=.657$] and Duration [$F(2,26)=8.66$, $p=.001$, $\eta_p^2=.400$], but the interaction was not [$F(2,26)=1.25$, $p=.30$, $\eta_p^2=.088$]. This was because audiovisual JND's were larger than visual ones, and both increased at 500 ms. Pairwise comparisons showed that the AV JND was lower for 40 than 500 ms duration ($p=.001$).

For Group 4 (different duration, 3 visual conditions) and right first analysis, there was a significant main effect of Modality [$F(1,6)=11.3$, $p=.015$, $\eta_p^2=.654$] because audiovisual JND's were larger than visual ones (Fig. 3D). Duration [$F(2,12)=2.99$, $p=.17$, $\eta_p^2=.258$] and interaction [$F(2,12)=.464$, $p=.64$, $\eta_p^2=.0720$] were not significant. The analysis of the results expressed as vision first for AV provided the same pattern of significances. There was a significant main effect of Modality [$F(1,6)=12.3$, $p=.013$, $\eta_p^2=.672$] because audiovisual JND's were larger than visual ones. Duration [$F(2,12)=3.27$, $p=.074$, $\eta_p^2=.353$] and interaction [$F(2,12)=.815$, $p=.47$, $\eta_p^2=.120$] were not significant.

In summary, JND's were generally larger for audiovisual than unisensory stimuli. There were some effects of duration. For Group 1, auditory JND's decreased with duration. For Group 3, visual and AV JND's increased with duration. Even though AV JND's seemed to decrease with visual stimulus duration for Group 4, this did not reach significance due to large variability.

Figure 3. about here

Points of subjective simultaneity

The PSS's for auditory, visual and audiovisual stimuli as a function of stimulus duration for Groups 1-4 are shown in Figure 4. Here any findings in the right first analyses would mean a bias to preferentially respond with either left or right key, meanwhile statistically significant findings in the vision first analyses for AV PSS would mean that either a visual or auditory lead is needed for subjective simultaneity.

For Group 1 (same duration, 3 auditory conditions) and right first analysis, there was a significant main effect of Modality [$F(1,10)=5.64$, $p=.039$, $\eta_p^2=.361$], with lower values for AV (Fig. 4A). The main effect of Duration [$F(2,20)=2.03$, $p=.16$, $\eta_p^2=.169$] and interaction [$F(2,20)=.454$, $p=.64$, $\eta_p^2=.043$] were not significant. The analysis of the results expressed as vision first for AV showed no effect of Duration [$F(2,20)=1.23$, $p=.30$, $\eta_p^2=.108$].

For Group 2 (different duration, 3 auditory conditions) and right first analysis, there were no significant effects [Modality $F(1,13)=1.67$, $p=.22$, $\eta_p^2=.114$; Duration $F(2,26)=1.75$, $p=.84$, $\eta_p^2=.013$; interaction $F(2,26)=2.87$, $p=.075$, $\eta_p^2=.181$]. The analysis of the results expressed as vision first for AV showed no effect of Duration [$F(2,26)=2.28$, $p=.123$, $\eta_p^2=.149$] (Fig. 4B).

For Group 3 (same duration, 3 visual conditions) and right first analysis, there were no significant effects [Modality $F(1,13)=3.41$, $p=.088$, $\eta_p^2=.208$; Duration $F(2,26)=.005$, $p=.99$, $\eta_p^2<.001$; interaction $F(2,26)=.031$, $p=.97$, $\eta_p^2=.002$]. The analysis of the results expressed as vision first for AV showed a significant effect of Duration [$F(2,28)=9.97$, $p=.001$, $\eta_p^2=.416$] because the PSS became more positive with increasing duration (Fig. 4C). Pairwise comparisons showed that the AV visual lead in PSS was larger for 500 than 10 ms ($p=.008$), and marginally larger than for 40 ms ($p=.054$).

For Group 4 (different duration, 3 visual conditions) and right first analysis, the main effect of Modality was not significant [$F(1,6)=.001$, $p=.98$, $\eta_p^2<.001$]. Duration [$F(2,12)=8.56$, $p=.005$, $\eta_p^2=.588$] and interaction [$F(2,12)=11.5$, $p=.002$, $\eta_p^2=.657$] were significant (Fig. 4D). A simple ANOVA for each modality showed that the effect of Duration was significant for visual [$F(2,16)=17.3$, $p<.001$, $\eta_p^2=.633$] but not for audiovisual [$F(2,20)=.089$, $p=.92$, $\eta_p^2=.011$]. In the visual condition the PSS was more positive for 500 ms than shorter stimulus durations (pairwise comparisons 10 vs 500 $p=.005$; 40 vs 500 $p<.001$). This means that participants had a bias to respond “right” at the longest duration. The analysis of the results expressed as vision first for AV showed no effect of Duration [$F(2,18)=1.67$, $p=.23$, $\eta_p^2=.156$].

In summary, in the cross-modal condition, PSS's were generally independent of duration, except that for Group 3 visual lead increased with duration.

Figure 4. about here

Individual variation

The main pattern emerging from the analyses of average values is that audiovisual JND's were larger than unisensory ones, and that duration effects were not very prominent. It should be noted that the individual variation was huge, as shown in Figure 5. Audiovisual JND's varied greatly between participants, and even though they tended to be lower when unisensory JND's were lower, they were mostly larger than unisensory JND's (Fig. 5A-B). Unisensory auditory JND's tended to be larger than visual JND's (Fig. 5C). There was no relationship between AV JND's and AV PSS's (Fig. 5D), instead both varied greatly and independently. There was no regularity regarding stimulus duration in AV or unisensory situations (Fig. 5A-F). In sum, there was no clear pattern in the individual variation regarding stimulus duration. However, stimulus modality showed a regularity so that the AV JND's were higher than the unisensory ones for the majority of individuals.

Figure 5. about here

Discussion

Auditory, visual and audiovisual temporal order judgments were measured as a function of stimulus duration to investigate the relationship between unisensory and cross-modal performance. The effects of duration were complex. On one hand, when the stimuli to be compared had equal durations, increasing their duration resulted in some effects. JND's decreased with duration for auditory stimuli (Group 1), but they increased with duration for visual stimuli (Group 3). The AV conditions were identical for Groups 1 and 3, but still JND's and PSS's were not affected by duration for Group 1, meanwhile both were higher at the longest duration for Group 3. On the other hand, when the stimuli to be compared had different durations, there were no duration effects (except for one case of apparently spurious unisensory response bias).

Overall, the audiovisual JND's were independent of duration. This was similar with the finding of Boenke et al. (2009). PSS's were also independent of duration, except that a longer visual lead was required for 500 ms than 10 ms stimuli with equal durations. This is in agreement with Kuling et al. (2012), except that they found it only for stimuli with different durations. Regarding the question whether only stimulus onset is used in TOJ, this was supported to the extent that stimulus duration did not influence the judgments. This was generally the case. Our findings thus suggest that TOJ is mainly based on stimulus onsets, particularly if one stimulus is of a fixed, short duration. Together

with previous findings, the cumulative evidence suggests that stimulus duration has a negligible effect on audiovisual TOJ.

The current findings were, however, more equivocal as far as unisensory TOJ is concerned. This is because auditory JND's decreased and visual JND's increased with duration when the stimuli to be compared had the same duration, even though duration was irrelevant to the task. This implies that the perceived onset may be influenced by stimulus duration. It seems that for auditory stimuli of increasing duration, the onset estimation becomes more accurate, resulting in decreasing JND's. Conversely, for visual stimuli of increasing duration, the onset estimation may become more variable, resulting in increasing JND's. At this point a tentative conclusion is that unisensory JND's may be influenced by stimulus duration. In addition to these different effects of duration in audition and vision, auditory JND's tended to be generally slightly larger than the visual ones, giving support to the existence of modality-specific timing mechanisms.

Attention is always important in perception, and prior entry in time perception is a phenomenon, which should be considered here. Prior entry (see Spence, Shore & Klein, 2001, for a review) means that attended stimuli are prioritized, being processed more rapidly. Or sometimes there is a response bias, so that an individual's criterion is shifted towards preferring the attended stimulus. In TOJ, prior entry is demonstrated by shifts in the PSS, so that they are shifted towards the attended modality. That is, if vision is attended, it reaches consciousness sooner, so that PSS is shifted towards more auditory lead. In the current study prior entry would mean preference over auditory or visual stimuli, depending on the focus of attention. For example, since some groups measured more visual than auditory TOJ's, they may have adopted a visual attention bias. The only PSS effect was that a larger visual lead was needed at the longest duration in Group 3 with 3 visual conditions. This is opposite to the prior entry prediction. Altogether, the role of potential attention biases appears small in the current study.

The temporal discrimination ability was generally poorer for cross-modal than unisensory stimuli. The finding of less precise audiovisual than unisensory synchrony perception has been reported in some previous studies (Alais and Cass, 2010; Fujisaki et al., 2012; Fujisaki and Nishida, 2005, 2010; Virsu et al., 2008; Zampini et al., 2003). Multisensory enhancement occurs within a window of temporal integration. The JND is considered to define the limits of this integration window (e.g. Vroomen and Keetels, 2010). The principle is that at asynchronies smaller than the JND, the stimuli are merged into a joint percept and can produce multisensory enhancement. The JND determines the temporal discrimination ability, so that at asynchronies at or beyond the JND, the stimuli are reliably perceived as separate events. The smaller the JND, the better the discrimination ability and the narrower the integration window. Of course, these are probabilistic constructs, so that the JND / window width is chosen at a certain level of performance beyond which integration is more likely. The current results

1 suggest that the window of integration is narrower for unisensory than cross-modal stimuli. There
2 might be some advantages in this: The temporal discrimination ability remains high within each
3 sensory modality, meanwhile multisensory perception remains coherent in despite some temporal
4 jitter in the input, since for example the conduction velocities differ between audition and vision (e.g.
5 Vatakis and Spence, 2010). That is, sound travels more slowly than light, so that an auditory signal
6 arrives at the sensory organ later than a visual signal at the same distance. In addition to this physical
7 difference in conduction velocity, there are differences between sensory modalities in neural
8 conduction velocities. Therefore, unisensory temporal processing is inherently more precise,
9 meanwhile audiovisual signals vary more in time, requiring a wider window of integration.
10

11 The finding that audiovisual JND's were larger than unisensory ones provides evidence against a
12 single, amodal central timing process (Sternberg and Knoll, 1973). If temporal order judgments were
13 always determined by a single process, there should be no difference in JND's whether the stimuli
14 were presented within one sensory modality or cross-modally. Instead, these results are more
15 consistent with the currently prominent view that the internal clock is not a single entity, but that there
16 are also sensory-specific components in the timing mechanism (Merchant, Zarco, and Prado, 2008,
17 and a review by Buetti, 2011). The cross-modal comparison of temporal order seems to be
18 accomplished after the time stamps of the stimuli have been determined separately in each modality.
19 This is supported by evidence from brain evoked potential studies showing that performance in
20 auditory (Bernasconi, Manuel, Murray, and Spierer, 2011) and audiovisual (Kösem, Gramfort, and
21 van Wassenhove, 2014) TOJ is reflected in auditory cortex activity. The need to shift attention
22 between modalities is a factor which may contribute to poorer cross-modal discrimination ability, as
23 described in the elegant model by Fujisaki and co-workers (2010; 2012).
24

25 There was great individual variability in TOJ. Both unisensory and cross-modal JND's varied across
26 participants, and AV PSS's ranged between auditory and visual leads. It appears that this individual
27 variation, which is a well-known finding in temporal perception (e.g. Boenke et al., 2009; Grabot &
28 van Wassenhove, 2017; Mollon and Perkins, 1996; Stevenson et al., 2012), is far greater than any
29 effects of stimulus duration. Still, the supremacy of unisensory over cross-modal temporal
30 discrimination ability can be seen at an individual level. Further, lower audiovisual JND's tended to be
31 associated with lower unisensory JND's, suggesting that precision in these tasks was related. In
32 contrast, audiovisual JND and PSS were unrelated, suggesting that they are determined by different
33 mechanisms. This implies that at least two processes contribute to individual differences; one related
34 to the precision of the time discrimination, another to the estimation of the relative timing of auditory
35 and visual signals, respectively.
36

37 In conclusion, when studying cross-modal phenomena, measuring also unisensory performance may
38 be informative. In audiovisual temporal order judgments, the discrimination ability was poorer for
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

cross-modal than unisensory stimuli. Overall, temporal order judgments were little affected by stimulus duration. Together, these findings give further support to the view that timing estimates are first obtained for audition and vision, and then compared across modalities.

Ethical approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Conflict of interest: The authors declare that they have no conflict of interest.

References

- Alais, D., & Cass, J. (2010). Multisensory perceptual learning of temporal order: audiovisual learning transfers to vision but not audition. *PLoS One*, 5(6), e11283. doi:[10.1371/journal.pone.0011283](https://doi.org/10.1371/journal.pone.0011283)
- Babkoff, H., & Fostick, L. (2013). The role of tone duration in dichotic temporal order judgment. *Atten Percept Psychophys*, 75(4), 654-660. doi:[10.3758/s13414-013-0449-6](https://doi.org/10.3758/s13414-013-0449-6)
- Bernasconi, F., Manuel, A. L., Murray, M. M., & Spierer, L. (2011). Pre-stimulus beta oscillations within left posterior sylvian regions impact auditory temporal order judgment accuracy. *Int J Psychophysiol*, 79(2), 244-248. doi:[10.1016/j.ijpsycho.2010.10.017](https://doi.org/10.1016/j.ijpsycho.2010.10.017)
- Boenke, L. T., Deliano, M., & Ohl, F. W. (2009). Stimulus duration influences perceived simultaneity in audiovisual temporal-order judgment. *Exp Brain Res*, 198(2-3), 233-244. doi:[10.1007/s00221-009-1917-z](https://doi.org/10.1007/s00221-009-1917-z)
- Borucki, E., & Berg, B. G. (2017). Isolating spectral cues in amplitude and quasi-frequency modulation discrimination by reducing stimulus duration. *Hearing Research*, 348, 129-133.
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10, 433-436.
- Bueti, D. (2011). The sensory representation of time. *Front Integr Neurosci*, 5, 34. doi:[10.3389/fnint.2011.00034](https://doi.org/10.3389/fnint.2011.00034)
- Eg, R., & Behne, D. M. (2015). Perceived synchrony for realistic and dynamic audiovisual events. *Front Psychol*, 6, 736. doi:[10.3389/fpsyg.2015.00736](https://doi.org/10.3389/fpsyg.2015.00736)
- Fujisaki, W., Kitazawa, S., & Nishida, S. (2012). Multisensory timing. In B. E. Stein (Ed.), *The new handbook of multisensory processing*. Cambridge, MA, US: MIT Press.
- Fujisaki, W., & Nishida, S. (2005). Temporal frequency characteristics of synchrony-asynchrony discrimination of audio-visual signals. *Exp Brain Res*, 166(3-4), 455-464. doi:[10.1007/s00221-005-2385-8](https://doi.org/10.1007/s00221-005-2385-8)
- Fujisaki, W., & Nishida, S. (2010). A common perceptual temporal limit of binding synchronous inputs across different sensory attributes and modalities. *Proc Biol Sci*, 277(1692), 2281-2290. doi:[rspb.2010.0243](https://doi.org/10.1098/rspb.2010.0243)
- Glabot, L., & van Wassenhove, V. (2017). Time order as psychological bias. *Psychological Science*, 28(5), 670-678.
- Grondin, S. (2010). Timing and time perception: a review of recent behavioral and neuroscience findings and theoretical directions. *Atten Percept Psychophys*, 72(3), 561-582. doi:[10.3758/APP.72.3.561](https://doi.org/10.3758/APP.72.3.561)
- Grondin, S. (2014). Why studying intermodal duration discrimination matters. [[10.3389/fpsyg.2014.00628](https://doi.org/10.3389/fpsyg.2014.00628)]. *Frontiers in Psychology*, 5.
- Hirsh, I. J., & Sherrick, C. E., Jr. (1961). Perceived order in different sense modalities. *J Exp Psychol*, 62, 423-432.
- Jaśkowski, P. (1991). Perceived onset simultaneity of stimuli with unequal durations. *Perception*, 20(6), 715-726.
- Jaśkowski, P. (1992). Temporal-order judgment and reaction time for short and long stimuli. *Psychological Research*, 54, 141-145.
- Jaśkowski, P. (1993). Temporal-order judgment and reaction time to stimuli of different rise times. *Perception*, 22(8), 963-970.
- Kleiner, M., Brainard, D. H., & Pelli, D. G. (2007). What's new in Psychtoolbox-3. *Perception ECVF Abstract Supplement*, 36.
- Kösem, A., Gramfort, A., & van Wassenhove, V. (2014). Encoding of event timing in the phase of neural oscillations. *Neuroimage*, 92, 274-284. doi:[10.1016/j.neuroimage.2014.02.010](https://doi.org/10.1016/j.neuroimage.2014.02.010)
- Krueger Fister, J., Stevenson, R. A., Nidiffer, A. R., Barnett, Z. P., & Wallace, M. T. (2016). Stimulus intensity modulates multisensory temporal processing. *Neuropsychologia*, 88, 92-100. doi:[http://dx.doi.org/10.1016/j.neuropsychologia.2016.02.016](https://doi.org/10.1016/j.neuropsychologia.2016.02.016)
- Kuling, I. A., van Eijk, R. L., Juola, J. F., & Kohlrausch, A. (2012). Effects of stimulus duration on audio-visual synchrony perception. *Exp Brain Res*, 221(4), 403-412. doi:[10.1007/s00221-012-3182-9](https://doi.org/10.1007/s00221-012-3182-9)

- Merchant, H., Zarco, W., & Prado, L. (2008). Do we have a common mechanism for measuring time in the hundreds of millisecond range? Evidence from multiple-interval timing tasks. *J Neurophysiol*, 99(2), 939-949. doi:01225.2007 10.1152/jn.01225.2007
- Mollon, J. D., & Perkins, A. J. (1996). Errors of judgement at Greenwich in 1796. *Nature*, 380(6570), 101-102. doi:10.1038/380101a0
- Patterson, R. D., Nimmo-Smith, I., Weber, D. L., & Milroy, R. (1982). The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold. *The Journal of the Acoustical Society of America*, 72, 1788-1803.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10, 437-442.
- Penney, T. B., Gibbon, J., & Meck, W. H. (2000). Differential effects of auditory and visual signals on clock speed and temporal memory. *J Exp Psychol Hum Percept Perform*, 26(6), 1770-1787.
- Salmela, V., & Laurinen, P. I. (2009). Low-level features determine brightness in White's and Benary's illusions. *Vision Research*, 49, 682-690.
- Schimmel, O., & Kohlrausch, A. (2008). On the influence of interaural differences on temporal perception of noise bursts of different durations. *J Acoust Soc Am*, 123(2), 986-997. doi:10.1121/1.2821979
- Shi, Z., Church, R. M., & Meck, W. H. (2013). Bayesian optimization of time perception. *Trends Cogn Sci*, 17(11), 556-564.
- Spence, C., Shore, D. I., & Klein, R. M. (2001). Multisensory prior entry. *J Exp Psychol Gen*, 130(4), 799-832.
- Sternberg, S., & Knoll, R. L. (1973). The perception of temporal order: Fundamental issues and a general model In S. Kornblum (Ed.), *Attention and performance IV*, (pp. 629-685). New York: Academic Press.
- Stevenson, R. A., Zemtsov, R. K., & Wallace, M. T. (2012). Individual differences in the multisensory temporal binding window predict susceptibility to audiovisual illusions. *Journal of Experimental Psychology: Human Perception and Performance*, 38(6), 1517-1529.
- Tiippana, K., Näsänen, R., & Rovamo, J. (1994). Contrast matching of two-dimensional compound gratings. *Vision Research*, 34, 1157-1163.
- van Wassenhove, V. (2009). Minding time in an amodal representational space. *Philos Trans R Soc Lond B Biol Sci*, 364(1525), 1815-1830. doi:364/1525/1815 10.1098/rstb.2009.0023
- Vatakis, A., & Spence, C. (2010). Audiovisual temporal integration for complex speech, object-action, animal call, and musical stimuli. In M. J. Naumer & J. Kaiser (Eds.), *Multisensory object perception in the primate brain*: Springer.
- Virsu, V., Oksanen-Hennah, H., Vedenpää, A., Jaatinen, P., & Lahti-Nuuttila, P. (2008). Simultaneity learning in vision, audition, tactile sense and their cross-modal combinations. *Exp Brain Res*, 186(4), 525-537. doi:10.1007/s00221-007-1254-z
- Vroomen, J., & Keetels, M. (2010). Perception of intersensory synchrony: a tutorial review. *Atten Percept Psychophys*, 72(4), 871-884. doi:72/4/871 10.3758/APP.72.4.871
- Wichmann, F., & Hill, N. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics*, 63(8), 1293-1313.
- Zampini, M., Shore, D. I., & Spence, C. (2003). Audiovisual temporal order judgments. *Exp Brain Res*, 152(2), 198-210. doi:10.1007/s00221-003-1536-z

Table legends

Table 1. Summary of the experimental design and participant groups.

Figure captions

Fig. 1 An illustration of the experimental design. Three examples of various stimulus types, durations, SOA's and left-right locations of the stimuli. Time runs downwards and SOA's are marked with light grey bars. On the left, an auditory stimulus pair with 10 ms duration. In the centre, an audiovisual stimulus pair with 40 ms duration. On the right, a visual stimulus pair with 40 and 500 ms durations

Fig. 2 A scatter plot of the coefficient of determination (R^2) vs JND. Right first JND's for auditory (black), visual (dark grey) and audiovisual (light grey) stimuli, and vision first JND's for audiovisual stimuli (white)

Fig. 3 The effect of duration on JND. **A:** Group 1. Same duration, i.e. stimuli on each trial had equal durations. Three auditory and audiovisual conditions, as well as one visual condition were measured. **B:** Group 2. Different duration, i.e. the duration of the auditory stimulus was varied while the other stimulus was always 10 ms. Three auditory and audiovisual conditions, as well as one visual condition were measured. **C:** Group 3. Same duration, i.e. stimuli on each trial had equal durations. Three visual and audiovisual conditions, as well as one auditory condition were measured. **D:** Group 4. Different duration, i.e. the duration of the visual stimulus was varied while the other stimulus was always 10 ms. Three visual and audiovisual conditions, as well as one auditory condition were measured. Unisensory auditory (circles), unisensory visual (diamonds) and cross-modal audiovisual right first (inverted triangles) and vision first (squares) JND's are plotted. Error bars depict the standard error of the mean.

Fig. 4 The effect of duration on PSS. Other details as in Fig. 3.

Fig. 5 Scatter plots of data for all participants in the study. **A:** Scatter plot of AV and A JND's. **B:** Scatter plot of AV and V JND's. **C:** Scatter plot of A and V JND's for 10 ms stimuli (the only duration for which all participants had both A and V measurements). **D:** Scatter plot of AV JND's and AV PSS's. **E:** Scatter plot of A JND's and A PSS's. **F:** Scatter plot of V JND's and V PSS's. Stimulus durations were 10 (black), 40 (gray) and 500 (white) ms. AV conditions are expressed as vision first. The diagonal line indicates that unisensory and cross-modal JND's are equal in A-C. Data points

above the diagonal line mean that unisensory JND's are larger than cross-modal JND's in B-C. The vertical line indicates physical simultaneity ($PSS=0$) in D-F. Negative PSS values indicate auditory lead, and positive PSS values indicate visual lead in D.

Group	n	Duration	Number of conditions:			A + A durations (ms)	V + V durations (ms)	A + V durations (ms)
			Auditory	Visual	AV			
1	15	Same	3	1	3	10 10	10 10	10 10
						40 40		40 40
						500 500		500 500
2	15	Different,	3	1	3	10 10	10 10	10 10
		A varied				10 40		40 10
						10 500		500 10
3	19	Same	1	3	3	10 10	10 10	10 10
							40 40	40 40
							500 500	500 500
4	15	Different,	1	3	3	10 10	10 10	10 10
		V varied					10 40	10 40
							10 500	10 500

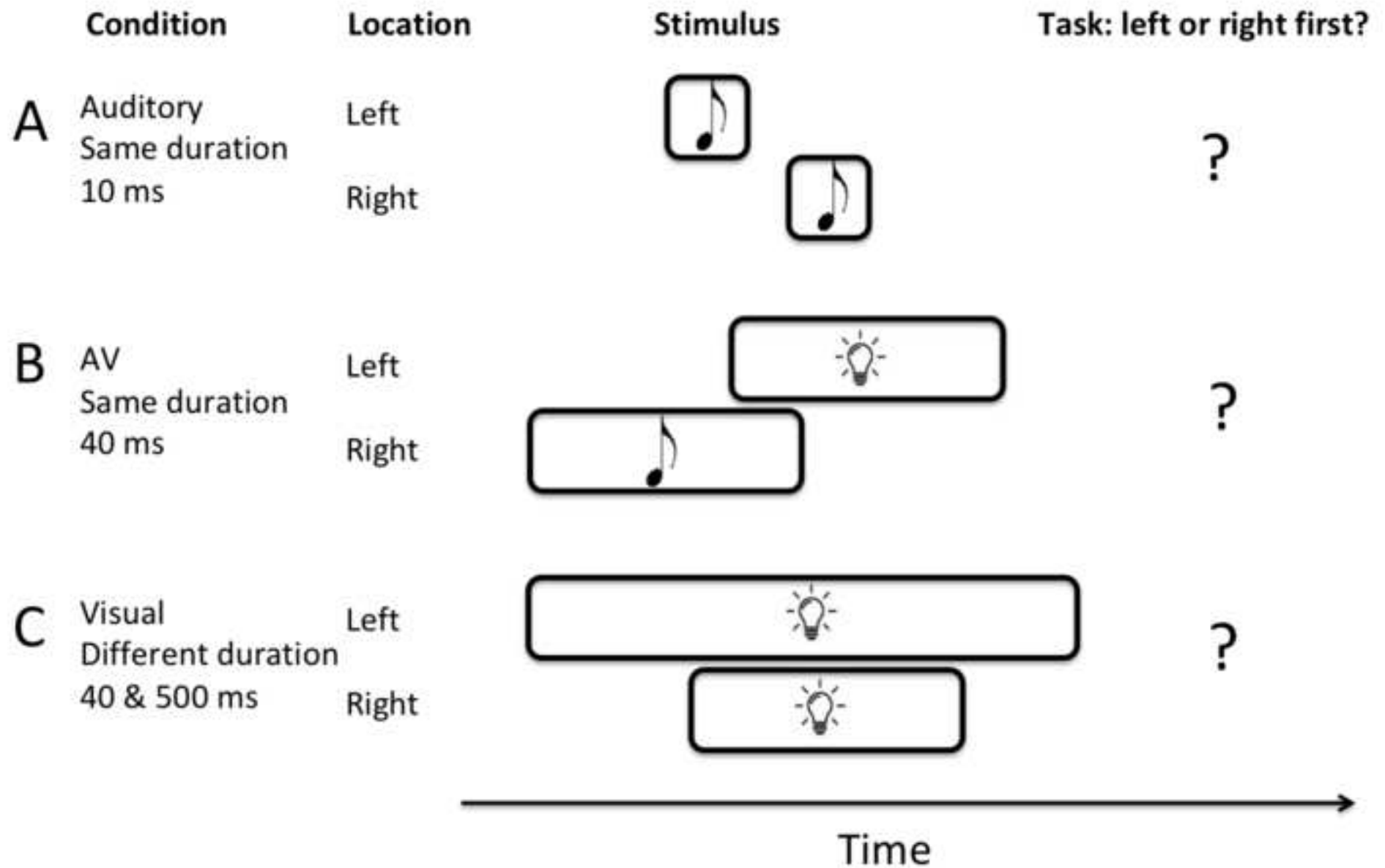


Figure 2

[Click here to download Figure](#)
[Previously_Fig2_Rsquare_JNDs_withoutliers_revised.tif](#)

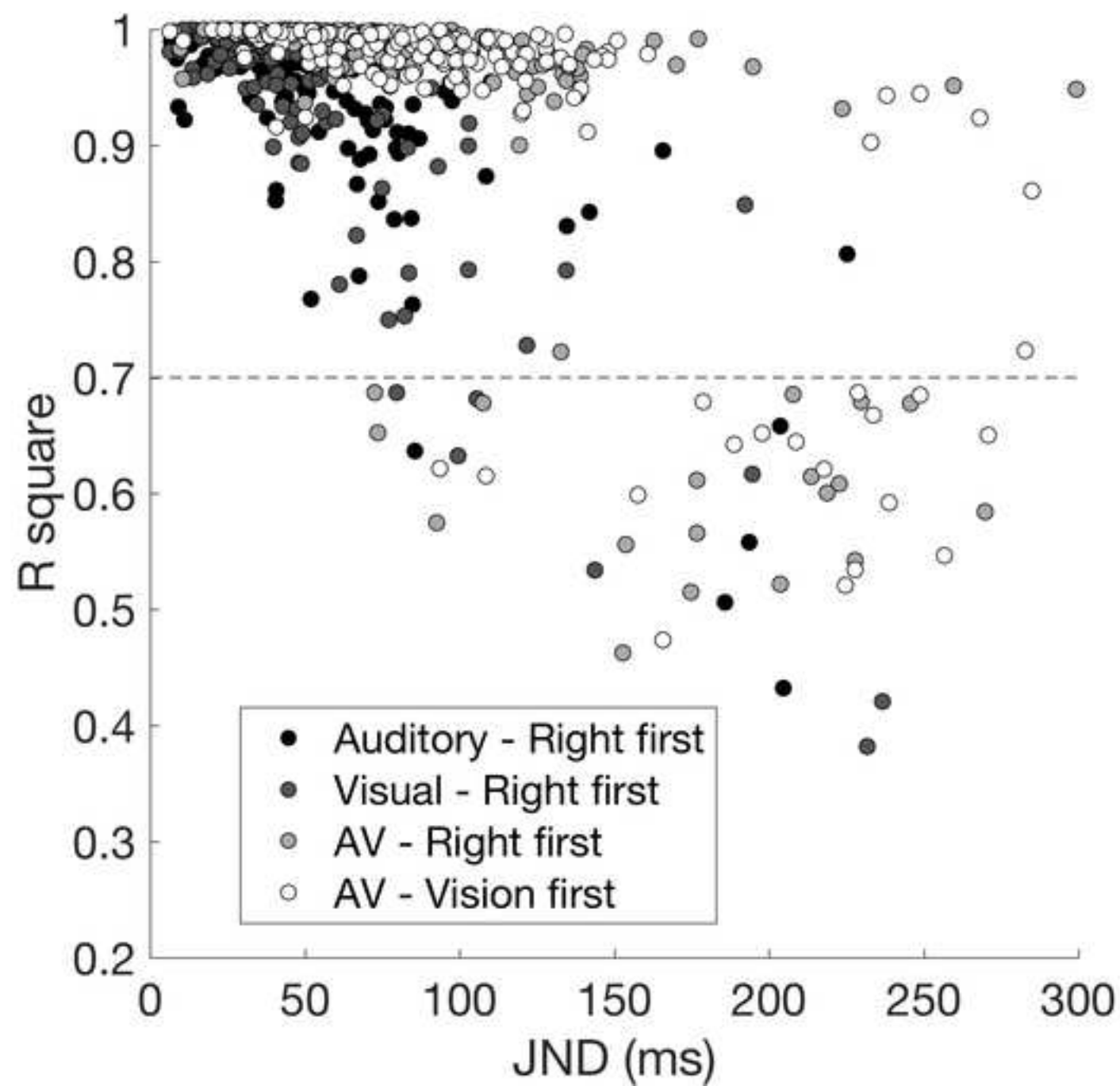


Figure 3

[Click here to download Figure Tiippana_Salmela_Fig1_JND_durations_revised.tif](#)

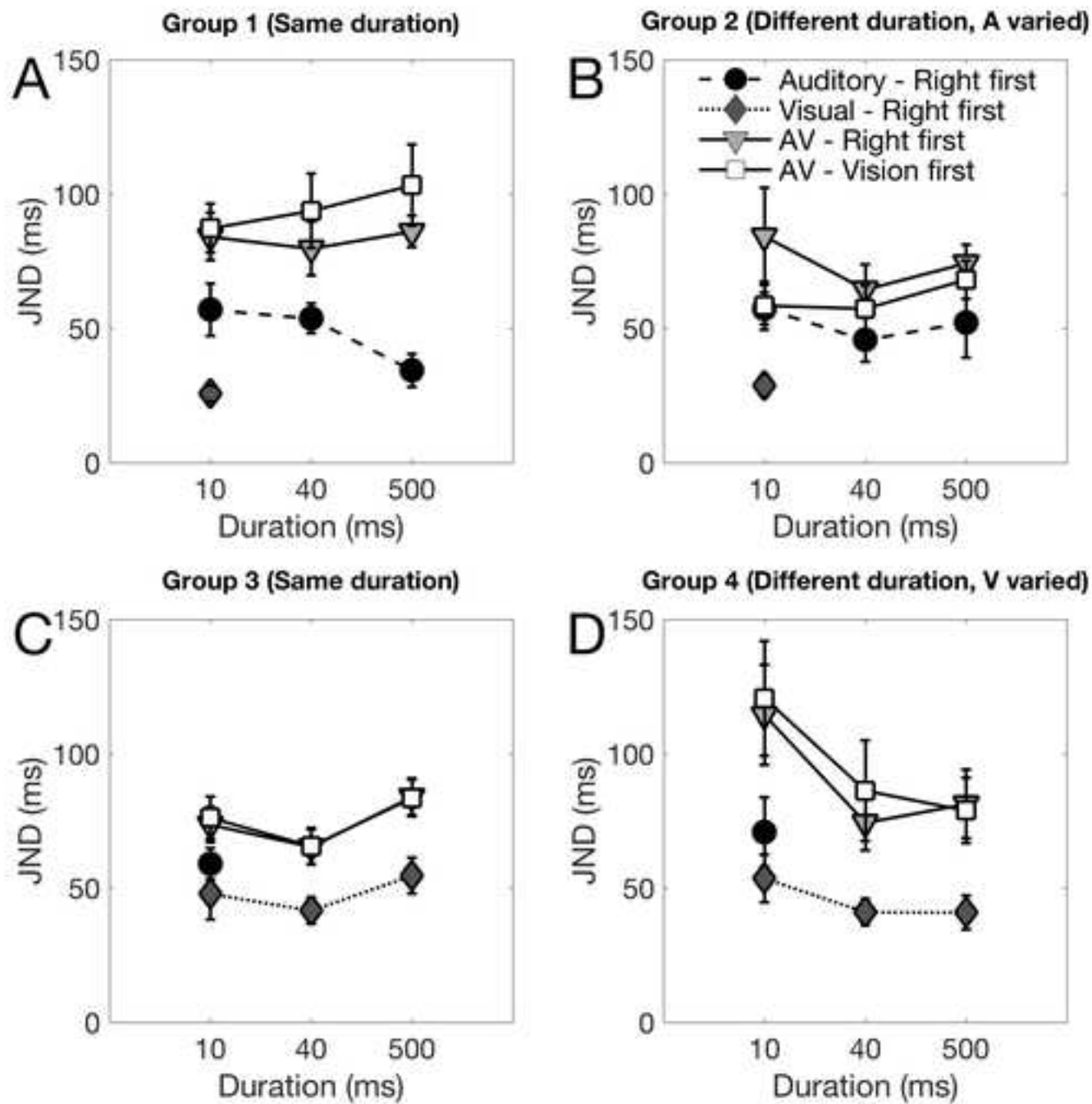


Figure 4

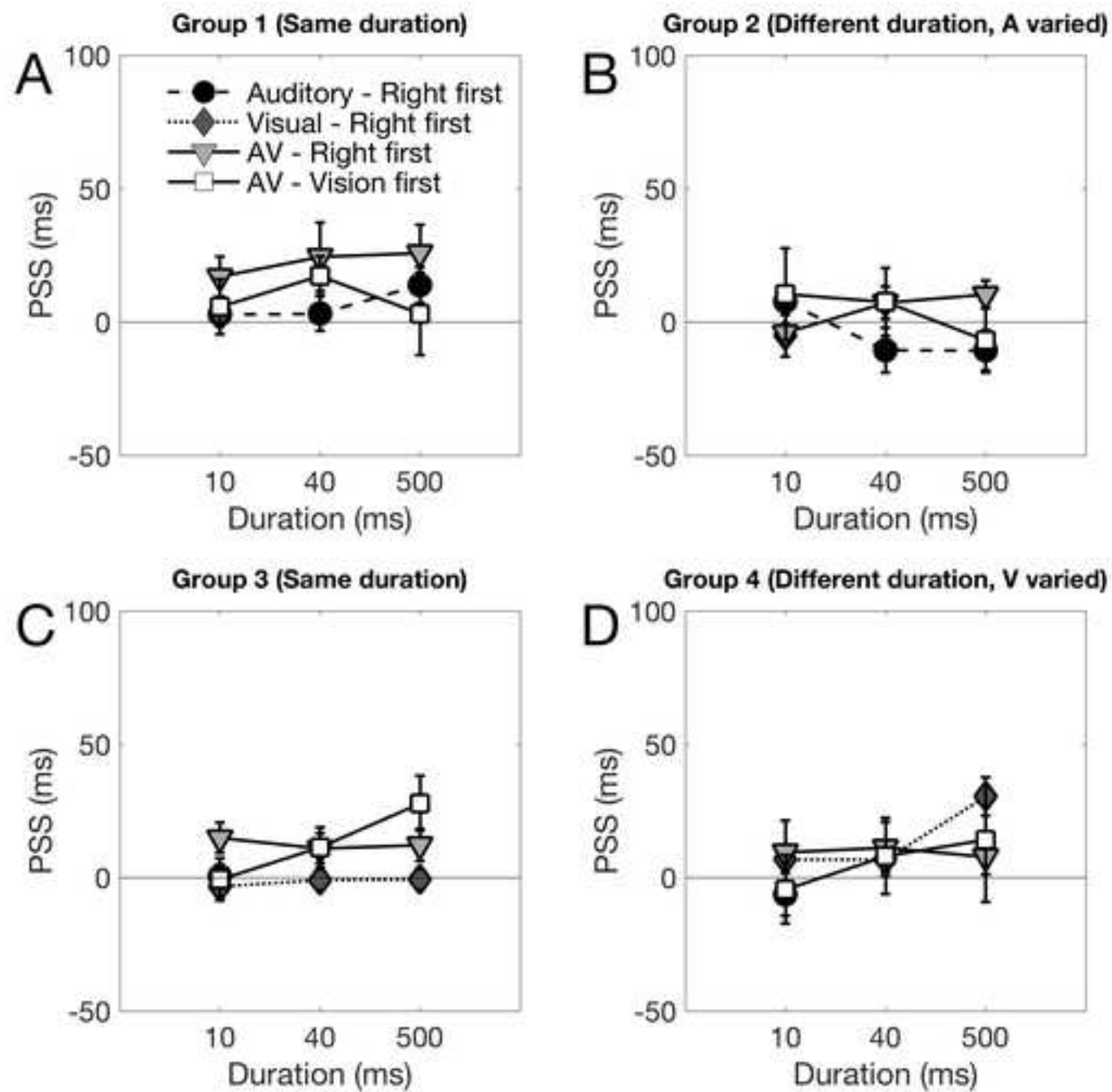


Figure 5

[Click here to download Figure Tiippana_Salmela_Fig3_scatterplots.tif](#)

